Paula Cristina Fartaria Marques

Licenciatura em Matemática

Licenciatura em Professores do Ensino Básico

O Teste de Esfericidade por Blocos de Matrizes para uma Amostra

Dissertação para obtenção do Grau de Mestre em

Matemática e Aplicações

Orientador: Filipe J. Marques, Doutor, FCT-UNL

Co-orientador: Carlos A. Coelho, Doutor, FCT-UNL

Júri:

Presidente: Prof. Doutor João Tiago Praça Nunes Mexia

Arguente: Prof. Doutor Luís Miguel Lindinho da Cunha Mendes Grilo

Vogal: Prof. Doutor Filipe José Gonçalves Pereira Marques

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2011

Paula Cristina Fartaria Marques

Licenciatura em Matemática

Licenciatura em Professores do Ensino Básico

O Teste de Esfericidade por Blocos de Matrizes para uma Amostra

Dissertação para obtenção do Grau de Mestre em

Matemática e Aplicações

Orientador: Filipe J. Marques, Doutor, FCT-UNL

Co-orientador: Carlos A. Coelho, Doutor, FCT-UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

O TESTE DE ESFERICIDADE POR BLOCOS DE MATRIZES PARA UMA AMOSTRA

Agradecimentos

Ao meu Orientador Professor Doutor Filipe J. Marques pela ajuda, paciência e apoio disponibilizado na realização deste trabalho.

Ao meu Co-orientador Professor Doutor Carlos A. Coelho pela sua compreensão e ajuda nas dificuldades sentidas.

Ao meu marido, pelo apoio e motivação nesta longa caminhada.

RESUMO

O objectivo do trabalho consiste em desenvolver distribuições quase-exactas para a estatística de razão de verosimilhanças do teste de esfericidade por blocos de matrizes para uma amostra, λ^*, extraída de uma população multivariada Normal. Através da decomposição da hipótese nula do teste em duas hipóteses nulas parciais, é possível obter não só a expressão da estatística de razão de verosimilhanças, como ainda a expressão do seu h-ésimo momento nulo e a função característica da variável aleatória $W = -\log\lambda^*$. Na primeira hipótese nula parcial é testada a independência de vários grupos de variáveis e na segunda a igualdade dos blocos diagonais da matriz de covariância. A decomposição da hipótese nula do teste em duas hipóteses nulas parciais, induz uma factorização na função característica de $W = -\log\lambda^*$ que serve de base para a construção das distribuições quase-exactas para W e para λ^*. Estas aproximações têm como distribuição uma distribuição Gama Quase-Inteira Generalizada (GQIG) ou mistura de duas ou três distribuições GQIG. Propomos ainda, para W, duas aproximações assimptóticas sob a forma de misturas de duas ou três distribuições Gama. Para avaliarmos a qualidade das aproximações assimptóticas e quase-exactas propostas são realizados estudos numéricos, que têm como base uma medida de proximidade entre distribuições, ela mesma baseada nas respectivas funções características. Os resultados obtidos permitem verificar a elevada precisão das distribuições quase-exactas e as suas boas propriedades assimptóticas.

Palavras-chave: Estatística de razão de verosimilhanças, distribuições quase-exactas, teste de independência, teste de igualdade de matrizes de covariância, Gama Inteira Generalizada, Gama Quase-Inteira Generalizada.

ABSTRACT

The aim of this study is to develop near-exact distributions for the likelihood ratio test statistic for the one sample block matrix sphericity test for a sample extracted from a multivariate normal population. Using the decomposition of the null hypothesis of the test in two partial null hypotheses, it is possible to obtain the expression of the likelihood ratio test statistic, the expression of its h-th null moment and the characteristic function of $W = -\log \lambda^*$. With the first null partial hypothesis we test the independence of k groups of variables and with the second one we test the equality of the k diagonal blocks of the covariance matrix. The decomposition of the null hypothesis in two null partial hypotheses, induces a factorization on the characteristic function $W= -\log \lambda^*$, which enables us to obtain near-exact distributions for W and λ^*. These approximations have Generalized Near-Integer Gamma distributions (GNIG) or mixtures of two or three distributions GNIG. We also propose two asymptotic approximations in the form of two or three mixtures of Gamma distributions. To test the quality of asymptotic and near-exact approximations, we carry out numerical studies, using a measure based on the respective characteristic functions, where we can check the high precision of near-exact distributions and its good asymptotic properties.

Keywords: Likelihood ratio statistic, near-exact distributions, test of independence, test of equal covariance matrices, Generalized Integer Gamma, Generalized Near-Integer Gamma.

ÍNDICE

LISTA DE TABELAS

INTRODUÇÃO

As estatísticas de teste mais comuns em estatística multivariada têm distribuições muito complexas e quase impossíveis de usar na prática sendo por isso necessário trabalhar com aproximações às suas actuais e complicadas distribuições.

Neste trabalho pretendemos obter aproximações para a distribuição da estatística usada no teste de esfericidade por blocos de matrizes para uma amostra. Chao & Gupta (1991) obtêm a estatística de razão de verosimilhanças do teste de esfericidade por blocos de matrizes, a expressão do seu h-ésimo momento e a sua distribuição exacta usando a expansão assimptótica de Barnes para a função Gama. Cardeño & Nagar (2001) apresentam a expressão do h-ésimo momento nulo da estatística de teste e obtêm para o caso de apenas dois blocos diagonais a expressão da densidade, usando a transformada inversa de Mellin e a definição de função de Meijer. Contudo, estes resultados são ainda difíceis de usar em termos práticos devido às complexas expressões das distribuições. Deste modo, propomos como aproximações para a distribuição da estatística do teste de esfericidade por blocos de matrizes para uma amostra, dois tipos de aproximações; as primeiras são aproximações assimptóticas baseadas em misturas de duas ou três distribuições Gama e obtidas através da técnica de acertar momentos e as segundas são aproximações quase-exactas. As distribuições quase-exactas são obtidas a partir de uma decomposição da hipótese nula do teste em duas hipóteses nulas parciais. Esta decomposição induz uma factorização na função característica do logaritmo da estatística de teste. As distribuições quase-exactas são então obtidas aproximando assimptoticamente uma parte da função característica pela função característica de uma distribuição Gama ou da mistura de duas ou três distribuições Gama. A substituição é feita de forma que a função característica resultante corresponda a uma distribuição conhecida e que possa ser usada na aplicação prática do teste através de cálculo de quantis e p-values. As distribuições quase-exactas, desenvolvidas neste trabalho, têm por base as distribuições Gama Inteira Generalizada (Coelho, 1998) e Gama Quase-Inteira Generalizada (Coelho, 2004). Para avaliar a qualidade das distribuições propostas vamos ainda considerar a aproximação apresentada em Chao & Gupta (1991), baseada no método desenvolvido por Box (1949).

No Capítulo 1, fazemos um resumo das diferentes distribuições que serão utilizadas ao longo deste trabalho, das quais destacamos a distribuição Gama Inteira Generalizada e a Gama Quase-Inteira Generalizada que são a base das distribuições quase-exactas propostas para a estatística de razão de verosimilhanças do teste.

No Capítulo 2, mostramos que podemos dividir a hipótese nula do teste, em duas hipóteses nulas parciais, uma para testar a independência de k grupos de p^* variáveis e outra para testar a igualdade de k matrizes de covariância. Com base nesta decomposição derivamos as expressões da estatística de razão de verosimilhanças, o *h*-ésimo momento nulo e também a função característica do logaritmo da estatística de razão de verosimilhanças. Será com base nesta expressão que vamos desenvolver distribuições quase-exactas para estatística de teste.

No Capítulo 3, propomos aproximações assimptóticas baseadas em misturas de 2 ou 3 distribuições Gama e desenvolvemos distribuições quase-exactas com base na factorização induzida pela decomposição da hipótese nula e que terão a distribuição de uma Gama Quase-Inteira Generalizada ou da mistura de duas ou três distribuições Gama Quase-Inteira Generalizadas.

Por último, no capítulo 4, usamos uma medida baseada nas fórmulas de inversão das funções características que é um “upper bound” para o módulo da diferença entre as respectivas distribuições, que permite avaliar a qualidade das várias aproximações.

CAPÍTULO 1 – CONCEITOS BÁSICOS

1.1 INTRODUÇÃO

Neste capítulo apresentamos as distribuições univariadas e multivariadas usadas nos capítulos que se seguem, assim como algumas das suas propriedades. Nos casos das distribuições mais comuns esta apresentação é feita essencialmente como objectivo de estabelecer a notação utilizada. Das distribuições apresentadas destacamos as distribuições Gama Inteira Generalizada (GIG) e Gama Quase-Inteira Generalizada (GQIG) que serão as distribuições base das aproximações quase-exactas desenvolvidas neste trabalho.

1.2 DISTRIBUIÇÃO GAMA

Dizemos que a v.a X segue uma distribuição Gama, se a função densidade for dada por,

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1},\ x \in (0, +\infty)$$

em que, $r > 0$ é o parâmetro de forma, $\lambda > 0$ é o parâmetro de taxa e $\Gamma(.)$ representa a usual função gama. Podemos escrever que $X \sim \Gamma(r, \lambda)$. A função característica da variável aleatória X é dada por,

$$\Phi_X(t) = \lambda^r (\lambda - it)^{-r},\ t \in R,$$

onde $i = (-1)^{1/2}$.

O h-ésimo momento da variável aleatória X com distribuição Gama de parâmetros r e λ é

$$E\left[X^h\right] = \frac{\Gamma(r+h)}{\Gamma(r)} \lambda^{-h},\ (h > -r).$$

1.3 DISTRIBUIÇÃO BETA

Dizemos que X é uma variável aleatória com distribuição Beta de parâmetros $\alpha > 0$ e $\beta > 0$, que se representa por $X \sim Beta(\alpha, \beta)$, se tiver função densidade de probabilidade dada pela expressão seguinte,

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (0 < x < 1),$$

onde $B(\alpha, \beta))$ é a função Beta.

A função característica de uma variável aleatória com distribuição $Beta(\alpha,\beta)$ é dada pela seguinte expressão,

$$\Phi_X(t) = {}_1F_1(\alpha;\alpha+\beta;it),\ t\in R \text{ e } i=(-1)^{1/2},$$

em que ${}_1F_1(\alpha;\beta;z)$ é a função hipergeométrica de Kummer (Abramowitz e Stegun, 1974), representada por,

$${}_1F_1(\alpha,\beta,z) = \frac{\Gamma(\beta)}{\Gamma(\alpha)}\sum_{j=0}^{\infty}\frac{\Gamma(\alpha+j)}{\Gamma(\beta+j)}\frac{z^j}{j!}.$$

O *h*-ésimo momento da variável aleatória X é dado por,

$$E\left(X^h\right) = \frac{B(\alpha+h,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+h)}{\Gamma(\alpha+\beta+h)} \quad , \quad (h>-\alpha).$$

1.4 DISTRIBUIÇÃO LOGBETA

Se $X \sim Beta(\alpha,\beta)$ com $\alpha>0$ e $\beta>0$, então a variável aleatória $Y=-\log X$ tem distribuição Logbeta com os parâmetros $\alpha>0$ e $\beta>0$ (Johnson et al., 1995). Esta pode ser representada por $Y \sim Logbeta(\alpha,\beta)$, sendo a sua função densidade de probabilidade dada por,

$$f_Y(y) = \frac{1}{B(\alpha,\beta)}e^{-\alpha y}\left(1-e^{-y}\right)^{\beta-1} \quad , \quad (y>0).$$

A função característica da variável aleatória $Y \sim Logbeta(\alpha,\beta)$ é representada da seguinte forma,

$$\begin{aligned}\Phi_Y(t) &= E\left(e^{itY}\right) = E\left(e^{-itlogX}\right) = E\left(X^{-it}\right)\\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha-it)}{\Gamma(\alpha+\beta-it)},\ (t\in R).\end{aligned}$$

1.5 A DISTRIBUIÇÃO GIG (GAMA INTEIRA GENERALIZADA)

Sejam,

$$X_j \sim \Gamma(r_j,\lambda_j) \qquad j=1,\dots,p$$

p variáveis aleatórias, independentes de taxa $\lambda_j>0$, com $\lambda_j \neq \lambda_{j'}$, quaisquer que sejam $j,j' \in \{1,\dots,p\}$, com $j \neq j'$.

Dizemos que a variável aleatória

$$Y = \sum_{j=1}^{p} X_j$$

tem distribuição GIG (Gama Inteira Generalizada) de profundidade p com parâmetros de forma r_j e parâmetros de taxa λ_j, $j = 1,\dots,p$ e representamos por,

$$Y \sim GIG(r_1,\dots,r_p;\lambda_1,\dots,\lambda_p;p).$$

As funções densidade e distribuição de Y (Coelho, 1998) são dadas respectivamente por,

$$f^{GIG}(y\,|\,r_1,\dots,r_p;\lambda_1,\dots,\lambda_p;p) = K\sum_{j=1}^{p} P_j(y)e^{-\lambda_j y} \quad , \quad (y>0)$$

e

$$F^{GIG}(y\,|\,r_1,\dots,r_p;\lambda_1,\dots,\lambda_p;p) = 1-K\sum_{j=1}^{p} P_j^*(y)e^{-\lambda_j y} \quad , \quad (y>0)$$

onde,

$$K = \prod_{j=1}^{p} \lambda_j^{r_j} \quad \text{e} \quad P_j(y) = \sum_{k=1}^{r_j} c_{j,k}\, y^{k-1}$$

e

$$P_j^*(y) = \sum_{k=1}^{r_j} c_{j,k}(k-1)!\sum_{i=0}^{k-1}\frac{y^i}{i!\lambda_j^{k-i}}$$

com

$$c_{j,r_j} = \frac{1}{(r_j-1)!}\prod_{\substack{i=1\\ i\neq j}}^{p}(\lambda_i-\lambda_j)^{-r_i}, \quad j = 1,\dots,p, \tag{1.5.1}$$

e

$$c_{j,r_j-k} = \frac{1}{k}\sum_{i=1}^{k}\frac{(r_j-k+i-1)!}{(r_j-k-1)!}R(i,j,p)c_{j,r_j-(k-i)}, \quad k = 1,\dots,r_j-1; \;\; j = 1,\dots,p \tag{1.5.2}$$

onde,

$$R(i,j,p) = \sum_{\substack{k=1\\ k\neq j}}^{p} r_k(\lambda_j-\lambda_k)^{-i} \qquad i = 1,\dots r_j - 1 \;. \tag{1.5.3}$$

1.6 A DISTRIBUIÇÃO GQIG (GAMA QUASE-INTEIRA GENERALIZADA)

A distribuição GQIG (Gama Quase-Inteira Generalizada) de profundidade $p+1$ é a distribuição da variável,

$$Z = Y_1 + Y_2$$

onde Y_1 e Y_2 são variáveis aleatórias independentes, a variável aleatória Y_1 com distribuição GIG de profundidade p, $Y_1 \sim GIG(r_1,...,r_p;\lambda_1,...,\lambda_p;p)$ e a variável aleatória Y_2 com distribuição Gama com parâmetro de forma r não inteiro e parâmetro de taxa $\lambda \neq \lambda_j$, $j = 1,...,p$ e denotamos por $Y_2 \sim \Gamma(r,\lambda)$.

A função densidade de Z é dada por,

$$f^{GQIG}(z \mid r_1,....,r_p, r;\lambda_1,....,\lambda_p, \lambda; p+1) = K\lambda^r \sum_{j=1}^{p} e^{-\lambda_j z} \sum_{k=1}^{r_j} \left\{ \mathrm{c}_{\mathrm{j,k}} \frac{\Gamma(k)}{\Gamma(k+r)} z^{k+r-1} {}_1F_1\left(r, k+r, -(\lambda-\lambda_j)z\right) \right\} \quad ,(z>0)$$

e a função distribuição por,

$$F^{GQIG}(z \mid r_1,....,r_p, r;\lambda_1,....,\lambda_p, \lambda; p+1) = \frac{\lambda^r z^r}{\Gamma(r+1)} {}_1F_1(r, r+1, -\lambda z)$$
$$- K\lambda^r \sum_{j=1}^{p} e^{-\lambda_j z} \sum_{i=1}^{r-1} \frac{\lambda_j^i z^{r+i}}{\Gamma(r+1+i)} {}_1F_1\left(r, r+1+i, -(\lambda-\lambda_j)z\right) \quad ,(z>0)$$

onde,

$$c^*_{j,k} = \frac{c_{j,k}}{\lambda_j^k}\Gamma(k)$$

com $c_{j,k}$ dados por (1.5.1) até (1.5.3).

1.7 DISTRIBUIÇÃO NORMAL MULTIVARIADA

Dizemos que $\underline{X} = [X_1,...,X_p]^T$ tem distribuição Normal *p*-Multivariada, facto que denotamos por, $\underline{X} \sim N_p(\underline{\mu},\Sigma)$, onde $\underline{\mu} = [\mu_1,...,\mu_p]^T$ e $\Sigma = [\mathrm{cov}(\underline{X}_i,\underline{X}_j)]$ com $i,j = 1,...,p$, se a função densidade de $\underline{X}$ for dada por,

$$f(\underline{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} . e^{-\frac{1}{2}(\underline{X}-\underline{\mu})^T \Sigma^{-1} (\underline{X}-\underline{\mu})} .$$

1.8 DISTRIBUIÇÃO WISHART

Esta distribuição é uma generalização da distribuição Qui-Quadrado. É uma distribuição de probabilidade conjunta de variáveis aleatórias representadas sob a forma matricial.

Sejam $\underline{X}_{\mathrm{i}} \sim N_p\left(\underline{0}, \Sigma\right)$, $i = 1, \ldots, n$ uma amostra aleatória. Então a matriz

$$W = \sum_{i=1}^{n} \underline{X}_i \, \underline{X}_i^T$$

tem uma distribuição de Wishart com matriz de parâmetro Σ e n graus de liberdade e denotamos este facto por $W \sim W_p(\Sigma, n)$.

CAPÍTULO 2 - O TESTE DE ESFERICIDADE POR BLOCOS DE MATRIZES PARA UMA AMOSTRA

2.1 INTRODUÇÃO

O teste de esfericidade usual permite testar se a hipótese nula de variáveis independentes, com igual variância é válida. O teste de esfericidade por blocos de matrizes é uma generalização do teste de esfericidade usual. Pretendemos, agora, testar se k grupos de variáveis aleatórias são independentes e têm iguais matrizes de covariância.

Consideremos uma amostra de dimensão N extraída de uma população Multivariada Normal, $N_p(\underline{\mu}, \Sigma)$.

Vamos testar a hipótese nula,

$$H_0 : \Sigma = \begin{bmatrix} \Delta & 0 & \dots & 0 \\ 0 & \Delta & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \Delta \end{bmatrix} \left(= \mathrm{I_k} \otimes \Delta\right), \text{ (Δ não especificada)} \tag{2.1.1}$$

onde as matrizes Δ são de ordem p^*, com $p = kp^*$.

2.2 A DECOMPOSIÇÃO DA HIPÓTESE NULA DO TESTE

Em Coelho & Marques (2009) a sugerida decomposição de uma hipótese, em hipóteses nulas parciais, revela-se uma boa opção quando se pretende obter distribuições quase-exactas para a estatística do teste. A hipótese nula em (2.1.1) pode ser decomposta em duas hipóteses nulas parciais, mais precisamente

$$H_o = H_{ob|oa} \; o \; H_{oa} \tag{2.2.1}$$

onde, para

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2k} \\ \vdots & & & \\ \Sigma_{k1} & \Sigma_{k2} & \dots & \Sigma_{kk} \end{bmatrix} \tag{2.2.2}$$

temos

$$H_{oa} : \Sigma_{ij} = 0 \quad \text{para } i \neq j \;, \quad (i, j = 1, \ldots k) \tag{2.2.3}$$

a hipótese nula para testar a independência dos k grupos de p^* variáveis e

$$H_{ob|oa} : \Sigma_{11} = \Sigma_{22} = ... = \Sigma_{kk} (= \Delta), \quad (\Delta \text{ não especificada}) \tag{2.2.4}$$

a hipótese nula para testar a igualdade das k matrizes de covariância de ordem p^*, assumindo H_{oa} como verdadeira.

2.3 A ESTATÍSTICA DE RAZÃO DE VEROSIMILHANÇAS E A EXPRESSÃO DO SEU *h*-ÉSIMO MOMENTO

A estatística modificada de razão de verosimilhanças para testar a hipótese nula em (2.2.3) é dada por,

$$\lambda_a^* = \frac{|A|^{n/2}}{\prod_{j=1}^{k} |A_{jj}|^{n/2}} \tag{2.3.1}$$

onde $n = N - 1$, $A = \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T$ e A_{jj} é a j-ésima matriz diagonal de ordem p^* de A.

A estatística modificada de razão de verosimilhanças para testar a hipótese nula em (2.2.4) é dada por,

$$\lambda_{b|a}^* = \frac{(kn)^{knp^*/2} \prod_{j=1}^{k} |A_{jj}|^{n/2}}{\prod_{j=1}^{k} n^{p^* n/2} |A^*|^{nk/2}} \tag{2.3.2}$$

onde $A^* = A_{11} + ... + A_{kk}$ e tem distribuição Wishart com nk graus de liberdade, o que é fácil de verificar uma vez que A_{jj} é a j-ésima matriz da diagonal de A de ordem p^*, as matrizes $A_{11}, A_{22}, ..., A_{kk}$ são independentes sob H_o e

$$A_{ii} \sim W_{p^*}(\Delta, n), \; i = 1, ..., k.$$

Atendendo às propriedades da distribuição de Wishart, referidas em Anderson (2003) temos

$$A^* = A_{11} + A_{22} + ... + A_{kk} \sim W_{p^*}(\Delta, \underbrace{n + n + ... + n}_{k \text{ vezes}})$$

ou seja,

$$A^* \sim W_{p^*}(\Delta, nk).$$

Segundo Anderson (2003) e tendo em conta a decomposição em (2.2.1), se λ_a^* , dada em (2.3.1), é a estatística de razão de verosimilhanças para testar H_{oa} e $\lambda_{b|a}^*$, dada em (2.3.2) é a estatística de razão de verosimilhanças para testar $H_{ob|oa}$, então $\lambda^* = \lambda_a^* \times \lambda_{b|a}^*$ é a estatística de razão de verosimilhanças para testar H_o. Portanto, a estatística modificada de razão de verosimilhanças é dada por

$$\lambda^* = \lambda_a^* \times \lambda_{b|a}^* = \frac{|A|^{n/2}}{\prod_{j=1}^{k} |A_{jj}|^{n/2}} \times \frac{(kn)^{knp^*/2} \prod_{j=1}^{k} |A_{jj}|^{n/2}}{\prod_{j=1}^{k} n^{p^*n/2} |A*|^{nk/2}} = \frac{|A|^{n/2} (kn)^{knp^*/2} \prod_{j=1}^{k} |A_{jj}|^{n/2}}{\prod_{j=1}^{k} |A_{jj}|^{n/2} \prod_{j=1}^{k} n^{p^*n/2} |A*|^{nk/2}}$$

$$= \frac{|A|^{n/2} (kn)^{knp^*/2}}{\prod_{j=1}^{k} n^{p^*n/2} |A*|^{nk/2}} = \frac{|A|^{n/2} (kn)^{knp^*/2}}{n^{knp^*/2} |A*|^{nk/2}} = \frac{|A|^{n/2} k^{knp^*/2}}{\left| k^{nk-nk/2} \sum_{j=1}^{k} A_{jj} \right|^{nk/2}}$$

$$= \frac{|A|^{n/2} k^{knp^*/2}}{k^{knp^*/2} \left| k^{-1} \sum_{j=1}^{k} A_{jj} \right|^{nk/2}} = \frac{|A|^{n/2}}{\left| \frac{1}{k} \sum_{j=1}^{k} A_{jj} \right|^{nk/2}} . \qquad (2.3.3)$$

Dada a independência das estatísticas de razão de verosimilhanças λ_a^* e $\lambda_{b|a}^*$ sob H_o, a expressão do h-ésimo momento nulo da estatística de teste,λ^* , pode ser obtida como produto das expressões dos h-ésimos momentos das estatísticas de razão de verosimilhanças, λ_a^* e $\lambda_{b|a}^*$ (para as expressões dos momentos veja-se Muirhead (1982) e Anderson (2003)), ou seja,

$$E\left[(\lambda *)^h\right] = E\left[\left(\lambda_{b|a}^* \times \lambda_a^*\right)^h\right] = E\left[\left(\lambda_{b|a}^*\right)^h \times \left(\lambda_a^*\right)^h\right]$$

$$= E\left[\left(\frac{(kn)^{knp^*/2} \prod_{j=1}^{k} |A_{jj}|^{n/2}}{\prod_{j=1}^{k} n^{np^*/2} |A*|^{nk/2}}\right)^h\right] \times E\left[\left(\frac{|A|^{n/2}}{\prod_{j=1}^{k} |A_{jj}|^{n/2}}\right)^h\right] =$$

$$= k^{knp^*h/2} \times \prod_{j=1}^{k} \frac{\Gamma_{p^*}\left(\frac{n}{2}(1+h)\right)}{\Gamma_{p^*}\left(\frac{n}{2}\right)} \times \frac{\Gamma_{p^*}\left(\frac{nk}{2}\right)}{\Gamma_{p^*}\left(\frac{nk}{2}(1+h)\right)} \times \frac{\Gamma_{kp^*}\left(\frac{1}{2}n + \frac{1}{2}nh\right)}{\Gamma_{kp^*}\left(\frac{1}{2}n\right)} \prod_{j=1}^{k} \frac{\Gamma_{p^*}\left(\frac{n}{2}\right)}{\Gamma_{p^*}\left(\frac{n}{2}(1+h)\right)} \qquad (2.3.4)$$

$$= k^{knp^*h/2} \frac{\Gamma_{p^*}\left(\frac{1}{2}nk\right)}{\Gamma_{p^*}\left(\frac{1}{2}(nk+nkh)\right)} \frac{\Gamma_{kp^*}\left(\frac{1}{2}(n+nh)\right)}{\Gamma_{kp^*}\left(\frac{1}{2}n\right)}$$

onde $\Gamma_p(.)$ é a função gama multivariada definida por

$$\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left(t - \frac{1}{2}(j-1)\right),$$

podemos a título de exemplo, indicar que,

$$\Gamma_{kp^*}\left(\frac{1}{2}(n+nh)\right) = \pi^{kp^*(kp^*-1)/4} \prod_{j=1}^{kp^*} \Gamma\left(\frac{1}{2}(n+nh) - \frac{1}{2}(j-1)\right) =$$

$$= \pi^{kp^*(kp^*-1)/4} \prod_{j=1}^{kp^*} \Gamma\left(\frac{1}{2}(n+nh-j+1)\right).$$

Portanto,

$$\mathrm{E}\left[(\lambda^*)^h\right] = \mathrm{k}^{\frac{1}{2}p^*knh} \prod_{j=1}^{p^*} \frac{\Gamma\left(\frac{1}{2}(nk-j+1)\right)}{\Gamma\left(\frac{1}{2}(nk+nkh-j+1)\right)} \times \prod_{j=1}^{kp^*} \frac{\Gamma\left(\frac{1}{2}(n+nh-j+1)\right)}{\Gamma\left(\frac{1}{2}(n-j+1)\right)}, \quad h > \frac{p-1}{n} \text{-1} .$$

2.4 A FUNÇÃO CARACTERÍSTICA DE W= - $\log\lambda^*$

Vamos usar a expressão do *h*-ésimo momento nulo de λ^* para obtermos a expressão da função característica de $W = -\log\lambda^*$ da seguinte forma:

$$\Phi_W(t) = E\left(e^{itW}\right) = E\left(\left(\lambda^*\right)^{-it}\right)$$

$$= \mathrm{k}^{-\frac{1}{2}p^*knit} \prod_{j=1}^{p^*k} \frac{\Gamma\left(\frac{1}{2}(n-nit-j+1)\right)}{\Gamma\left(\frac{1}{2}(n-j+1)\right)} \times \prod_{j=1}^{p^*} \frac{\Gamma\left(\frac{1}{2}(nk-j+1)\right)}{\Gamma\left(\frac{1}{2}(nk-nkit-j+1)\right)}, \ t \in R. \quad (2.4.1)$$

A decomposição da hipótese nula em (2.2.1) induz uma factorização na função característica de W. Usando a factorização em (2.3.4) podemos escrever a função característica em (2.4.1) como o produto das funções características de $W_1 = -\log\lambda_a^*$ e de $W_2 = -\log\lambda_{b|a}^*$.

Assim,

$$\Phi_W(t) = \Phi_{W_1}(t) \times \Phi_{W_2}(t),$$

em que

$$\Phi_{W_1}(t)=E\left(e^{-itW_1}\right)=\frac{\Gamma_{kp^*}\left(\frac{1}{2}n-\frac{itn}{2}\right)}{\Gamma_{kp^*}\left(\frac{1}{2}n\right)}\prod_{j=1}^{k}\frac{\Gamma_{p^*}\left(\frac{1}{2}n\right)}{\Gamma_{p^*}\left(\frac{1}{2}n-\frac{itn}{2}\right)}$$

e

$$\Phi_{W_2}(t)=E\left(e^{-itW_2}\right)=$$

$$=\frac{(kn)^{-itknp^*/2}}{\prod_{j=1}^{k}n^{-itnp^*/2}}\frac{\Gamma_{p^*}\left(\frac{nk}{2}\right)}{\Gamma_{p^*}\left(\frac{nk}{2}-\frac{itnk}{2}\right)}\prod_{j=1}^{k}\frac{\Gamma_{p^*}\left(\frac{n}{2}-\frac{itn}{2}\right)}{\Gamma_{p^*}\left(\frac{n}{2}\right)}$$

Então, obtemos

$$\Phi_W(t)=\underbrace{\frac{\Gamma_{p^*k}\left(\frac{1}{2}n-\frac{1}{2}itn\right)}{\Gamma_{p^*k}\left(\frac{1}{2}n\right)}\prod_{j=1}^{k}\frac{\Gamma_{p^*}\left(\frac{n}{2}\right)}{\Gamma_{p^*}\left(\frac{n}{2}(1-it)\right)}}_{\Phi_{W_1}(t)}\times$$

$$\times\underbrace{\frac{(nk)^{-nkp^*it/2}\,\Gamma_{p^*}\left(\frac{nk}{2}\right)}{\prod_{j=1}^{k}n^{-p^*nit/2}\,\Gamma_{p^*}\left(\frac{nk}{2}(1-it)\right)}\prod_{j=1}^{k}\frac{\Gamma_{p^*}\left(\frac{n}{2}(1-it)\right)}{\Gamma_{p^*}\left(\frac{n}{2}\right)}}_{\Phi_{W_2}(t)} \qquad (2.4.2)$$

2.5 FACTORIZAÇÃO DAS FUNÇÕES CARACTERÍSTICAS DE W_1 E W_2

Com o objectivo final de desenvolver distribuições quase-exactas para a estatística modificada de razão de verosimilhanças, λ^*, vamos utilizar factorizações das funções características de $W_1=-\log\lambda_a^*$ e $W_2=-\log\lambda_{b|a}^*$. Estas factorizações mostram que as distribuições exactas de W_1 e W_2 podem ser representadas sob a forma da soma de duas variáveis aleatórias independentes, uma com distribuição GIG e outra com distribuição correspondente à soma de variáveis independentes com distribuição Logbeta multiplicadas, eventualmente, por uma constante.

2.5.1 A FUNÇÃO CARACTERÍSTICA DE W_1= - logλ_a^*

Em Coelho (2004), o autor apresenta uma possível factorização para a função característica de $W_1 = -\log \lambda_a^*$, na seguinte forma,

$$\Phi_{W_1}(t) = \prod_{j=1}^{p-2} \left(\frac{n-p+j}{2} \right)^{z_j} \times \left(\frac{n-p+j}{2} - it\frac{n}{2} \right)^{-z_j} \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}$$

podemos reescrever a função característica anterior da seguinte forma:

$$\Phi_{W_1}(t) = \prod_{j=1}^{p-2} \left(\frac{\frac{n-p+j}{2}}{\frac{n-p+j-itn}{2}} \right)^{z_j} \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}$$

$$= \prod_{j=1}^{p-2} \left(\frac{\frac{n-p+j}{n}}{\frac{n-p+j-itn}{n}} \right)^{z_j} \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}$$

$$= \prod_{j=1}^{p-2} \left(\frac{n-p+j}{n} \right)^{z_j} \left(\frac{n-p+j}{n} - it \right)^{-z_j} \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}$$

$$= \prod_{j=1}^{p-2} \left(\frac{n-(p-j)}{n} \right)^{z_j} \left(\frac{n-(p-j)}{n} - it \right)^{-z_j} \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}$$

$$= \underbrace{\prod_{j=2}^{p\text{-}1} \left(\frac{n-j}{n} \right)^{z_{p-j}} \left(\frac{n-j}{n} - it \right)^{-z_{p-j}}}_{\Phi_{1,1}(t)} \qquad (2.5.1.1)$$

$$\underbrace{\times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}-\frac{n}{2}it\right)}{\Gamma\left(\frac{n}{2}-\frac{n}{2}it\right)\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)} \right\}^{m*}}_{\Phi_{1,2}(t)} \qquad (2.5.1.2)$$

com $$m^* = \begin{cases} 0\,, & p^* \text{ par} \\ k\,, & p^* \text{ ímpar}\,. \end{cases} \tag{2.5.1.3}$$

Sendo os parâmetros z_j dados por,

$$z_j = \begin{cases} 0\,, & j=2 \\ h_1 - m^*, & j=3 \\ h_2 + m^*, & j=4 \\ z_{j-2} + h_{j-2}\,, & j=5,\dots, p \end{cases} \tag{2.5.1.4}$$

com

$$h_j = \left(p_i \ \left(i=1,\dots,k\right) \geq j\right) - 1,\ \ j=1,\dots p-2 \tag{2.5.1.5}$$

onde k é o número de grupos de variáveis com um número ímpar de variáveis.

A função característica $\Phi_{1,1}(t)$ corresponde à soma de $p-2$ variáveis aleatórias com distribuição Gama com parâmetros de forma z_j, inteiros, dados em (2.5.1.4) e taxas $\frac{n-j}{n}$ com $j=2,\dots,p-1$, isto é, uma distribuição GIG (Coelho, 1998) de profundidade $p-2$.

A função característica $\Phi_{1,2}(t)$ corresponde à soma de m^* variáveis aleatórias independentes com distribuição Logbeta multiplicadas por $\frac{n}{2}$ com parâmetros $\frac{n}{2}-\frac{1}{2}$ e $\frac{n}{2}$.

2.5.2 A FUNÇÃO CARACTERÍSTICA DE $W_2 = -\log\lambda_{b|a}^*$

Em Coelho & Marques (2011) os autores obtêm a seguinte factorização para $\Phi_{W_2}(t)$,

$$\Phi_{W_2}(t) = \underbrace{\prod_{k=1}^{p-1} \left(\frac{n-k}{n}\right)^{r_k} \left(\frac{n-k}{n} - it\right)^{r_k}}_{\Phi_{2,1}(t)} \tag{2.5.2.1}$$

$$\underbrace{\begin{aligned}&\times \prod_{j=1}^{\lfloor p/2 \rfloor} \prod_{k=1}^{q} \frac{\Gamma(a_j + b_{jk})\Gamma(a_j + b_{jk}^{*} - nit)}{\Gamma(a_j + b_{jk}^{*})\Gamma(a_j + b_{jk} - nit)} \\ &\qquad\qquad \times \left(\prod_{k=1}^{q} \frac{\Gamma(a_p + b_{pk})\Gamma\left(a_p + b_{pk}^{*} - \frac{n}{2}it\right)}{\Gamma(a_p + b_{pk}^{*})\Gamma\left(a_p + b_{pk} - \frac{n}{2}it\right)} \right)^{p \amalg 2}\end{aligned}}_{\Phi_{2,2}(t)} \qquad (2.5.2.2)$$

onde,

$$r_k = \begin{cases} r_k^{*}, & k = 1,\ldots,p\text{-}1 \text{ e } k \neq p\text{-}1\text{-}2\alpha_1 \\ \mathrm{r}_k^{*} + a^{*}, & k = p\text{-}1\text{-}2\alpha_1 \end{cases}, \qquad (2.5.2.3)$$

com,

$$a^{*} = (p \amalg 2)(\alpha_2 - \alpha_1)\left(q - \frac{p-1}{2} + q\left\lfloor \frac{p}{2q} \right\rfloor \right),$$

e,

$$p \amalg 2 = \left\lfloor \frac{p+1}{2} \right\rfloor - \left\lfloor \frac{p}{2} \right\rfloor = \begin{cases} 0, & p \text{ par} \\ 1, & \text{caso contrário}; \end{cases}$$

$$r_k^{*} = \begin{cases} c_k, & k \in \{1,\ldots,\alpha+1\} \\ q\left(\left\lfloor \frac{p}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor \right), & k \in \{\alpha+2,\ldots,\min(\mathrm{p}-2\alpha_1, p-1)\} \\ & k \in \{2+p-2\alpha_1,\ldots 2\lfloor p/2 \rfloor - 1;\ \text{passo } 2\} \\ q\left(\left\lfloor \frac{p+1}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor \right), & k \in \{1+p-2\alpha_1,\ldots,p-1;\ \text{passo } 2\} \end{cases} \qquad (2.5.2.4)$$

e onde para $k = 1,\ldots,\alpha$,

$$c_k = \left\lfloor \frac{q}{2} \right\rfloor \left((k-1)q - 2((q+1) \amalg 2)\left\lfloor \frac{q}{2} \right\rfloor \right) + \left\lfloor \frac{q}{2} \right\rfloor \left\lfloor \frac{q + k \amalg 2}{2} \right\rfloor$$

e

$$c_{\alpha+1} = -\left(\left\lfloor \frac{p}{2} \right\rfloor - \alpha \left\lfloor \frac{q}{2} \right\rfloor\right)^2 + q\left(\left\lfloor \frac{p}{2} \right\rfloor - \left\lfloor \frac{\alpha+1}{2} \right\rfloor\right)$$
$$+ (q \amalg 2)\left(\alpha \left\lfloor \frac{p}{2} \right\rfloor + \frac{\alpha \amalg 2}{4} + \frac{\alpha^2}{4} + \alpha^2 \frac{q}{2}\right). \qquad (2.5.2.5)$$

A função característica $\Phi_{2,1}(t)$ corresponde à soma de $p-1$ variáveis aleatórias independentes com distribuição Gama, com parâmetros de forma r_k , inteiros, dados em (2.5.2.3) e taxas $\frac{n-k}{n}$ com $k=1,\dots,p-1$, isto é, uma distribuição GIG de profundidade $p-1$.

A função característica $\Phi_{2,2}(t)$ corresponde à soma de $\lfloor p/2 \rfloor \times q + q \times p \amalg 2$ variáveis aleatórias independentes com distribuição Logbeta, as primeiras $\lfloor p/2 \rfloor \times q$ multiplicadas por n e as segundas $q \times p \amalg 2$ multiplicadas $\frac{n}{2}$.

CAPÍTULO 3 – DISTRIBUIÇÕES ASSIMPTÓTICAS E QUASE-EXACTAS PARA W E λ^*

3.1 INTRODUÇÃO

Neste capítulo desenvolvemos vários tipos de aproximações para a estatística λ^*. Começamos por apresentar uma aproximação para W apresentada em Chao & Gupta (1991), baseada no tradicional método de Box (1949). Em seguida propomos duas aproximações assimptóticas para W na forma de mistura de duas ou três distribuições Gama, com o mesmo parâmetro de taxa e que acertam os primeiros quatro ou seis momentos da distribuição exacta.
Finalmente, desenvolvemos distribuições quase-exactas para W e λ^* que têm a distribuição de uma GQIG ou da mistura de duas ou três distribuições GQIG que acertam os primeiros dois, quatro ou seis momentos da distribuição exacta.

3.2 APROXIMAÇÃO BASEADA NO MÉTODO DE BOX

Em Chao & Gupta (1991), embora os autores apresentem um teste mais geral que o teste de esfericidade por blocos de matrizes para uma amostra, na prática apenas estudam este teste porque assumem e não testam a igualdade de matrizes de covariância associadas ao número de populações multivariadas normais consideradas.
Assim, com as devidas correcções, podemos utilizar a aproximação apresentada pelos referidos autores, que é baseada no usual Método de Box, como uma aproximação assimptótica para $W = -\log \lambda^*$.
A aproximação é apresentada sob a forma de uma mistura de Qui-quadrados,

$$P(-2\rho \log \lambda^* \leq z) = P\left(\chi_f^2 \leq z\right) + \omega_2\left(P\left(\chi_{f+4}^2 \leq z\right) - P\left(\chi_f^2 \leq z\right)\right) + O\left(n^{-3}\right) \qquad (3.2.1)$$

onde,

$$\omega_2 = \frac{2}{3n^2\rho^2} \frac{pk(pk-1)(pk+1)(p+2)}{32} - \frac{p(p+1)(p+2)}{32k^2} - \frac{3n^2\left(1-\rho^2\right)f}{24}$$

com $f = \frac{p}{2}(k-1)[p(k+1)+1]$ graus de liberdade,

e,

$\rho = 1 - \frac{p}{12nfk}(k-1)\left[2p^2\left(k^2+1\right)(k+1) + 3p\left(k^2+k+1\right) - (k+1)\right]$, onde k é o numero de grupos e p é o número total de variáveis.

Atendendo à expressão (3.2.1) e a que $W = -\log \lambda^*$, obtemos

$$P(2\rho W \le z) \cong P\left(\chi_f^2 \le z\right) + \omega_2 P\left(\chi_{f+4}^2 \le z\right) - \omega_2 P\left(\chi_f^2 \le z\right)$$

e podemos considerar,

$$P(2\rho W \le z) \cong (1-\omega_2) P\left(\chi_f^2 \le z\right) + \omega_2 P\left(\chi_{f+4}^2 \le z\right).$$

Usando o facto de que,

$$\Phi_W(t) = E\left[e^{itW}\right] = E\left[e^{it(-\log\lambda^*)}\right] = E\left[e^{\frac{it}{2\rho}\left(-2\rho\log\lambda^*\right)}\right]$$

e sendo χ_f^2 um caso especial da distribuição Gama, pois

$$X \sim \chi_f^2 \Leftrightarrow X \sim \Gamma\left(\frac{f}{2}, \frac{1}{2}\right)$$

então,

$$\Phi_W(t) \cong (1-\omega_2)\left[\left(\frac{1}{2}\right)^{f/2}\left(\frac{1}{2} - \frac{it}{2\rho}\right)^{-f/2}\right] + \omega_2\left[\left(\frac{1}{2}\right)^{(f+4)/2}\left(\frac{1}{2} - \frac{it}{2\rho}\right)^{-(f+4)/2}\right] =$$

$$= (1-\omega_2)\left[\left(\frac{1}{2}\right)^{f/2}\left(\frac{1}{2\rho}\right)^{-f/2}(\rho - it)^{-f/2}\right] + \omega_2\left[\left(\frac{1}{2}\right)^{\left(f/2\right)+2}\left(\frac{1}{2\rho}\right)^{-\left(f/2\right)-2}(\rho - it)^{-\left(f/2\right)+2}\right] =$$

$$= (1-\omega_2)\left[\left(\frac{1/2}{1/2\rho}\right)^{f/2}(\rho - it)^{-f/2}\right] + \omega_2\left[\left(\frac{1/2}{1/2\rho}\right)^{\left(f/2\right)+2}(\rho - it)^{-\left(f/2\right)-2}\right].$$

Desta forma, obtemos uma aproximação para $\Phi_W(t)$, sob a forma de uma mistura de duas distribuições Gama,

$$\Phi_{Box}(t) = (1-\omega_2)(\rho)^{\frac{f}{2}}(\rho - it)^{-\frac{f}{2}} + \omega_2(\rho)^{\frac{f}{2}+2}(\rho - it)^{-\frac{f}{2}-2}. \tag{3.2.2}$$

3.3 APROXIMAÇÕES BASEADAS EM MOMENTOS

Considerando a aproximação obtida por Box (1949) propomos duas novas aproximações assimptóticas para W, sob a forma de duas misturas de distribuições Gama que acertam os quatro ou seis primeiros momentos da distribuição exacta. Nas diferentes misturas, à semelhança do que acontece na aproximação de Box, consideramos igual parâmetro de taxa para as diferentes Gamas na mistura.

Especificando, propomos como aproximações para $\Phi_W(t)$ a função característica da mistura de duas ou três distribuições Gama, ambas com o mesmo parâmetro de taxa,

$$\Phi_{M2G}(t)=\sum_{j=1}^{2} p_j \lambda^{s_j} (\lambda - it)^{-s_j} \ , \tag{3.3.1}$$

e

$$\Phi_{M3G}(t)=\sum_{j=1}^{3} p_j \lambda^{s_j} (\lambda - it)^{-s_j} \tag{3.3.2}$$

com $p_2 = 1 - p_1;\ p_3 = 1 - p_2 - p_1;\ p_j, s_j, \lambda > 0 \ \ j = 1,...,3.$

As expressões em (3.3.1) e (3.3.2), são obtidas pela resolução do seguinte sistema de equações:

$$i^h \sum_{j=1}^{k} p_j \frac{\Gamma(s_j + h)}{\Gamma(s_j)} \lambda^{-h} = \frac{\partial^h \Phi_{MkG}(t)}{\partial t^h}\Big|_{t=0} = \frac{\partial^h \Phi_W(t)}{\partial t^h}\Big|_{t=0} \ , \quad h = 1,...,2k.$$

tomando *k*=2, no caso da mistura de duas distribuições Gama, ou *k*=3, no caso da mistura de três distribuições Gama.

3.4 APROXIMAÇÕES QUASE-EXACTAS PARA W E λ^*

Considerando as factorizações apresentadas na secção 2.5, onde estão bem patentes as semelhanças entre as distribuições de W_1 e W_2 , podemos reescrever a função característica de $W = -\log \lambda *$, através do teorema seguinte:

Teorema 1: *A função característica de* $W = -\log \lambda *$ *pode ser apresentada da forma,*

$$\Phi_W(t) = \underbrace{\prod_{j=1}^{p-1} \left(\frac{n-j}{n}\right)^{v_j} \left(\frac{n-j}{n} - it\right)^{-v_j}}_{\Phi_{W_1^*}(t)} \times \underbrace{\Phi_{1,2}(t) \times \Phi_{2,2}(t)}_{\Phi_{W_2^*}(t)} \tag{3.4.1}$$

com $\Phi_{1,2}(t)$ e $\Phi_{2,2}(t)$ *dadas respectivamente em (2.5.1.2) e (2.5.2.2) e onde os* v_j *são dados por,*

$$v_j = \begin{cases} r_j & j = 1 \\ r_j + z_{p-j} & j = 2,..., p-1 \end{cases} \tag{3.4.2}$$

com *os* z_j *dados em (2.5.1.4) e os* r_j *dados em (2.5.2.3).*

Com base no teorema 1, desenvolvemos distribuições quase-exactas que têm por base a seguinte construção,

$$\Phi_W(t) = \underbrace{\Phi_{W_1^*}(t)}_{\textit{Dist. GIG}} \quad \times \quad \underbrace{\Phi_{W^*_2}(t)}_{\textit{soma de Logbeta ind.}}$$

substituição assimptótica para $\Phi_{W_2^*}(t)$

$$\underbrace{\Phi_W(t) \approx \Phi_{W_1^*}(t) \quad \times \quad \overbrace{\sum_{j=1}^{h/2} \theta_j \mu^{\delta_j} (\mu - it)^{-\delta_j}}}_{\textit{mistura de 2 ou 3 distribuições GQIG (h=4 ou h=6)}} \quad .$$

(sendo uma única distribuição GQIG para h=2)

Portanto,

$$\Phi_W(t) = \underbrace{\Phi_{W_1^*}(t)}_{\text{Distribuição GIG}} \times \tilde{\Phi}_{W_2^*}(t) \approx \Phi_W(t), \tag{3.4.3}$$

onde $\Phi_{W_1^*}(t)$ e $\Phi_{W_2^*}(t)$ são dadas em (3.4.1) e $\tilde{\Phi}_{W_2^*}(t)$ é a função característica em (3.4.5) utilizada para aproximar a função característica $\Phi_{W_2^*}(t)$.

Propomos que a função característica $\tilde{\Phi}_{W_2^*}(t)$ seja a função característica de uma distribuição Gama ou da mistura de duas ou três distribuições Gama, dependente do número de momentos que se pretende acertar. Assim, as funções características $\Phi_{W_2^*}(t)$ e $\tilde{\Phi}_{W_2^*}(t)$ têm as primeiras duas, quatro ou seis derivadas em zero iguais, por outras palavras temos,

$$\frac{d^j}{dt^j} \Phi_{W_2^*}(t)|_{t=0} = \frac{d^j}{dt^j} \tilde{\Phi}_{W_2^*}(t)|_{t=0} \ , \ j = 1,\dots,h \tag{3.4.4}$$

para h=2, 4 ou 6 consoante $\tilde{\Phi}_{W_2^*}(t)$ seja a função característica de uma distribuição Gama ou a função característica da mistura de duas distribuições Gama ou a função característica da mistura de três distribuições Gama com os mesmos parâmetros de taxa, isto é,

$$\tilde{\Phi}_{W_2^*}(t) = \sum_{j=1}^{h/2} \theta_j \mu^{\delta_j} (\mu - it)^{-\delta_j} \tag{3.4.5}$$

com pesos $\theta_j > 0 \ \left(j = 1,\dots, h/2 \ \text{ com } \ h = 2, 4 \text{ ou } 6 \right)$ e $\sum_{j=1}^{h/2} \theta_j = 1$.

Desta forma, podemos escrever a função característica quase-exacta do logaritmo da estatística de razão de verosimilhanças, $W = -\log \lambda^*$, na forma (3.4.3) onde $\tilde{\Phi}_{W_2^*}(t)$ é a função característica de uma distribuição Gama, da mistura de duas distribuições Gama ou da mistura de três distribuições Gama, obtendo como distribuições quase-exactas, respectivamente, uma distribuição GQIG ou a mistura de duas ou três distribuições GQIG de profundidade p.

No teorema que se segue, são apresentadas as distribuições e as funções densidade de probabilidade para as distribuições quase exactas de λ^*.

Teorema 2: *As distribuições quase-exactas para $W = -\log \lambda^*$ são uma GQIG ou a mistura de duas ou três distribuições GQIG de profundidade p para h=2,4 ou 6. As funções densidade de probabilidade das distribuições quase-exactas para λ^* são dadas por (veja-se secções 1.5 e 1.6),*

$$\sum_{v=1}^{h/2} \theta_v f^{GQIG}\left(-\log \omega \,|\, v_1,\ldots,v_{p-1},\delta_v; \frac{n-1}{n},\ldots,\frac{n-p+1}{n},\mu; p\right)\frac{1}{\omega}$$

e as funções distribuição das distribuições quase-exactas para λ^* são dadas por

$$1-\sum_{v=1}^{h/2} \theta_v F^{GQIG}\left(-\log \omega \,|\, v_1,\ldots,v_{p-1},\delta_v; \frac{n-1}{n},\ldots,\frac{n-p+1}{n},\mu; p\right)$$

com os v_j dados em (3.4.2), e onde, para h=2

$$\mu = \frac{m_1}{m_2 - m_1^2} \qquad \text{e} \qquad \delta_1 = \frac{m_1^2}{m_2 - m_1^2}$$

com

$$m_j = i^{-j} \frac{\partial^j}{\partial t^j} \Phi_{W_2^*}(t)|_{t=0}\,, \quad j = 1,2$$

e para h=4 ou 6 os valores dos parâmetros θ_v, δ_v e μ são obtidos como solução numérica do sistema de equações em (3.4.4), isto é,

$$\frac{d^j}{dt^j}\Phi_{W_2^*}(t)|_{t=0} = \frac{d^j}{dt^j}\tilde{\Phi}_{W_2^*}(t)|_{t=0} \quad , j = 1,\ldots.,h$$

com

$$p_{h/2} = 1-\sum_{i=1}^{h/2-1} p_i\,.$$

Na próxima secção podemos constatar que estas distribuições quase-exactas fornecem excelentes aproximações para W e para λ^*, e são mais fáceis de usar em termos práticos o que facilita o cálculo de quantis quase-exactos e *p*-values.

CAPÍTULO 4 – ESTUDOS NUMÉRICOS

Para avaliarmos a qualidade das aproximações desenvolvidas no capítulo anterior, usamos uma medida de proximidade entre as funções características. Esta já utilizada em diversos trabalhos para avaliar qualidade de distribuições quase-exactas (Grilo & Coelho, 2007; Coelho & Marques, 2011).

Seja Y uma variável aleatória contínua com suporte S, em que $F_Y(y)$ é a sua função distribuição, $f_Y(y)$ é a sua função densidade de probabilidade e $\Phi_Y(t)$ é a função característica. Seja X_n uma variável aleatória, em que $F_{X_n}(y)$, $f_{X_n}(y)$ e $\Phi_{X_n}(t)$ são, respectivamente, a função distribuição, a função densidade e a função característica da variável aleatória X_n.

A medida é dada por,

$$\Delta = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\Phi_Y(t)-\Phi_{X_n}(t)}{t}\right|dt\,, \tag{4.1}$$

com

$$\max_{y\in S}\left|F_Y(y)-F_{X_n}(y)\right|\leq\Delta\,.$$

Nas tabelas seguintes vamos denotar por GQIG, M2GQIG e M3GQIG respectivamente as distribuições quase-exactas GQIG, mistura de duas distribuições GQIG e mistura de três distribuições GQIG. Além disso, denotemos por M2G e M3G a mistura de duas ou três distribuições Gama que acertam quatro ou seis momentos da distribuição exacta e que correspondem às aproximações propostas na secção 3.3. Finalmente denotaremos por Box a aproximação assimptótica obtida em Chao & Gupta (1991).

Nas tabelas 4.1 e 4.2 apresentamos os valores da medida Δ, para o caso em que p^* está fixo e p, k e n aumentam.

Tabela 4.1 – Valores da medida Δ para *p*=2* e valores crescentes de *p*, *k* e *n*.

p	p^*	k	n	*GQIG*	*M2GQIG*	*M3GQIG*	*M2G*	*M3G*	*BOX*
6	2	3	8	4.0×10^{-6}	6.0×10^{-9}	3.0×10^{-11}	4.7×10^{-4}	6.4×10^{-5}	1.9×10^{-2}
8	2	4	10	1.5×10^{-6}	3.7×10^{-10}	1.8×10^{-12}	5.9×10^{-4}	9.2×10^{-5}	4.7×10^{-2}
10	2	5	12	4.4×10^{-7}	1.2×10^{-10}	7.5×10^{-14}	1.7×10^{-4}	1.9×10^{-5}	4.0×10^{-2}
14	2	7	16	1.0×10^{-7}	1.5×10^{-11}	4.5×10^{-15}	7.5×10^{-4}	1.4×10^{-4}	1.8×10^{-1}
18	2	9	20	2.7×10^{-8}	1.9×10^{-12}	2.9×10^{-16}	1.5×10^{-4}	1.5×10^{-4}	3.1×10^{-1}

Tabela 4.2 – Valores da medida Δ para *p*=3* e valores crescentes de *p*, *k* e *n*.

p	p^*	*k*	*n*	*GQIG*	*M2GQIG*	*M3GQIG*	*M2G*	*M3G*	*BOX*
9	3	3	11	4.5×10^{-6}	3.6×10^{-9}	4.0×10^{-12}	6.2×10^{-4}	1.0×10^{-4}	5.8×10^{-2}
12	3	4	14	1.2×10^{-6}	4.3×10^{-10}	2.9×10^{-13}	7.1×10^{-4}	1.2×10^{-4}	1.2×10^{-1}
15	3	5	17	4.3×10^{-7}	8.2×10^{-11}	3.8×10^{-14}	7.6×10^{-4}	1.4×10^{-4}	2.1×10^{-1}
21	3	7	23	8.3×10^{-8}	1.7×10^{-12}	9.0×10^{-16}	8.0×10^{-4}	2.5×10^{-4}*	4.1×10^{-1}
27	3	9	29	2.5×10^{-8}	6.3×10^{-13}	5.2×10^{-17}	8.1×10^{-4}	3.7×10^{-4}*	6.4×10^{-1}

* Os parâmetros foram calculados com dificuldade na convergência.

Podemos observar que os valores de Δ apresentados pela distribuição quase-exacta GQIG são bem melhores do que os apresentados pelas distribuições assimptóticas , misturas de duas ou três Gamas. As misturas de duas ou três distribuições GQIG são as distribuições que apresentam os valores mais baixos, sendo a distribuição M3GQIG a que apresenta os valores mais baixos de todas. A aproximação assimptótica denotada por Box é a que apresenta os valores mais altos.

Nas tabelas 4.3 e 4.4 apresentamos o caso em que apenas o valor de n aumenta sucessivamente e os restantes parâmetros permanecem inalterados.

Tabela 4.3 - Valores da medida Δ para *p=8*, *p*=4*, *k=2* e valores crescentes de n

p	p^*	*k*	*n*	*GQIG*	*M2GQIG*	*M3GQIG*	*M2G*	*M3G*	*BOX*
8	4	2	10	3.1×10^{-6}	2.9×10^{-10}	9.1×10^{-13}	5.6×10^{-4}	8.5×10^{-5}	3.3×10^{-2}
8	4	2	50	2.1×10^{-7}	6.9×10^{-13}	1.8×10^{-14}	1.8×10^{-7}	7.3×10^{-10}	3.2×10^{-5}
8	4	2	100	5.5×10^{-8}	4.3×10^{-14}	3.2×10^{-18}	1.1×10^{-8}	2.7×10^{-11}	3.3×10^{-6}

Tabela 4.4 - Valores da medida Δ para *p=8*, *p*=4*, *k=2* e valores crescentes de n

p	p^*	*k*	*n*	*GQIG*	*M2GQIG*	*M3GQIG*	*M2G*	*M3G*	*BOX*
9	3	3	10	4.5×10^{-6}	3.6×10^{-9}	4.0×10^{-12}	6.2×10^{-4}	1.0×10^{-4}	5.8×10^{-2}
9	3	3	50	4.6×10^{-7}	2.7×10^{-11}	1.2×10^{-14}	2.6×10^{-7}	1.3×10^{-9}	9.0×10^{-5}
9	3	3	100	1.2×10^{-7}	2.0×10^{-12}	1.5×10^{-15}	1.6×10^{-8}	3.4×10^{-11}	9.2×10^{-6}

Analisando as tabelas 4.3 e 4.4, podemos concluir que os valores de Δ da distribuição quase-exacta GQIG só apresenta os melhores valores, comparada com a mistura de duas ou três Gamas, para $n=10$. No entanto, os valores da distribuição quase-exacta GQIG conseguem ser melhores do que os da aproximação assimptótica denotada por Box, mas ficam aquém dos valores Δ das misturas de duas ou três distribuições GQIG. Novamente, a mistura de três distribuições GQIG apresenta os valores mais baixos.

Em todas as tabelas verificamos as boas propriedades assimptóticas das distribuições quase-exactas para valores crescentes de *p* e *n*.

CONCLUSÃO

Neste trabalho procurámos contribuir para que as dificuldades levantadas na utilização do teste de esfericidade por blocos de matrizes para uma amostra, devidas essencialmente à estrutura complexa da distribuição exacta da estatística de razão de verosimilhanças, fossem de alguma forma ultrapassadas. Neste sentido propusemos para a distribuição da estatística de teste duas aproximações assimptóticas sob a forma de misturas de duas ou três distribuições Gama e desenvolvemos, com base numa decomposição da hipótese nula do teste, distribuições quase exactas que correspondem a uma distribuição Gama Quase-Inteira Generalizada ou a misturas destas distribuições.

A utilização de uma decomposição da hipótese nula, do teste de esfericidade por blocos de matrizes para uma amostra, em duas hipóteses nulas parciais, uma para estudar a independência dos grupos de variáveis e outra para estudar a igualdade das matrizes de covariância, permitiu obter de uma forma simples a expressão da estatística de razão de verosimilhanças, a expressão do seu h-ésimo momento nulo e ainda a expressão da função característica do logaritmo da estatística de razão de verosimilhanças.

A factorização induzida na função característica do logaritmo da estatística de teste pela decomposição da hipótese nula considerada, juntamente com os resultados obtidos em Coelho & Marques (2011) e Coelho (2004) permitiram o desenvolvimento das distribuições quase-exactas. As distribuições quase-exactas obtidas revelaram-se excelentes aproximações quando comparadas com as outras aproximações propostas e com a aproximação obtida em Chao & Gupta (1991) e têm distribuições conhecidas, com expressões fáceis de implementar e de usar em termos computacionais, o que as torna uma óptima ferramenta para usar na prática.

Verificámos, ainda, que as aproximações quase-exactas revelam boas propriedades assimptóticas não só para valores crescentes do tamanho da amostra, mas também para valores crescentes do número de variáveis e do número de grupos de variáveis.

BIBLIOGRAFIA

Abramowitz, M. & Stegum, I. A. (1974). *Handbook of Mathematical Functions*, 9rd ed., Dover, New York.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed., J. Wiley & Sons, New York.

Box, G. E. P.(1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.

Cardeño, L. & Nagar K. D. (2001). Testing Block Sphericity of a Covariance Matrix. *Divulgaciones Matemáticas*, 9, 25-34.

Chao, C. C.; Gupta A. K.(1991). Testing of Homogeneity of Diagonal Blocks with Blockwise Independence. *Communication in Statistics –Theory and Methods*, 20, 1957-1969.

Coelho, C. A. (1998).The Generalized Integer Gamma Distribution – A basis for distributions in Multivariate Statistics. *Journal of Multivariate Analysis*, 64, 86-102.

Coelho, C. A.(2004). The Generalized Near-Integer Gamma Distribution: a basis for 'near exact' approximations to the distributions of statistics which are the product of an odd number of independent Beta random variables. *Journal of Multivariate Analysis*, 89, 191-218.

Coelho, C. A. & Marques, F. J. (2011). Near-exact distributions for the likelihood ratio test statistic for testing the equality of several variance-covariance matrices in elliptically contourned distributions. *Computational Statistic*(in print).

Coelho, C. A. & Marques, F. J. (2009).The advantage of decomposing elaborate hypotheses on covariance matrices into conditionally independent hypotheses in building near-exact distributions for the test statistics. *Linear Algebra and Its Applications,* 430, 2592-2606.

Grilo L.M. & Coelho C.A. (2007). Development and Comparative Study of two Near-exact Approximations to the Distribution of the Product of an Odd Number of Independent Beta Random Variables. *Journal of Statistical Planning and Inference*, 137, 1560-1575.

Johnson, N.L. & Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions. 2rd ed.J.Wiley& Sons, New York.

Marques, F. J. & Coelho, C. A. (2011). A general near-exact distribution theory for the most common likelihood ratio test statistics used in Multivariate Statistics. *Test*,20, 180-203.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory*. 2rd ed. J. Wiley & Sons, New York.